Can street segments indexed for accessibility form the basis for delineating housing submarkets?

Housing Studies

Yang Xiao[a], Chris Webster[b*], Scott Orford[c]

a: Department of Urban Planning, Tongji University, 1239 Siping Road, Shanghai, 200092, China
b: Faculty of Architecture, The University of Hong Kong, 4/F, Knowles Building, Pokfulam Road, Hong Kong
c: School of Geography and Planning, Cardiff University, Glamorgan Building, King Edward VII Avenue, Cardiff, Wales, CF10 3WA

* corresponding author: Prof. Chris Webster, cwebster@hku.hk.

Abstract

We test an approach to spatial housing sub-market delineation using street segment as the spatial unit and using finely grained measures of accessibility derived from spatial network analysis. The underlying idea is that street segment connectivity captures fine variations in home-buyers' preferences for the location. The advantage of the approach is that it is spatially fine grained; it uses the street segment, intuitively the most fundamental spatial unit for spatial housing market analysis; it allows the use of statistical tests to optimize within-sub-market similarities, identifying spatial groups of street segments with the most similar accessibility features; it avoids the predefined arbitrary geographic boundaries usually used in spatial sub-market delineation; it increases the variability of accessibility information in sub-market delineation, accessibility being the principal spatial determinant of housing price; and it allows for normalized measures of accessibility at different spatial scales making it appropriate for comparative analysis across cities. Using a case study of Cardiff, UK, we compare the results with a market segmentation scheme based on prior-knowledge, notably one relying on building type classification. We conclude that street layout can be used to efficiently delineate housing submarkets, and that the estimation is very close to the scheme requiring prior-knowledge. It has advantages, however that make it worthy of further investigation, namely its adaptability, scale-specificity and lower reliance on local knowledge of housing market culture and data.

Keywords: housing submarket, street segments, accessibility, geographic boundary, network analysis

1.0 Introduction

Housing markets in a city, region or country are not uniform entities; they comprise distinct expressions of demand and supply spatially defined. For analytical, professional and policy purposes, the segmentation of housing submarkets should therefore be based on a clear understanding of spatially-specific demand and supply dynamics. There are many studies arguing for the existence of submarkets and suggesting methods of identifying them (Leishman, 2001, Watkins, 2001, Bramley et al., 2008, Jones et al., 2009, Leishman, 2009, Leishman et al., 2013, Bourassa et al., 1999, Park, 2013, Pryce, 2013, Goodman and Thibodeau, 2007).

Despite a large amount of research, there is no clear agreement about how spatial submarkets should best be identified for housing studies (Adair et al., 1996, Watkins, 2001, Bourassa et al., 1999, Leishman, 2009). In practice, the crucial questions of how to identify close substitutes and what level of aggregation is appropriate, are often answered in an ad hoc manner. Predefined geographical boundaries, often arbitrary in terms of housing market behaviour, are adopted as the spatial unit for submarket delineation (Goodman and Thibodeau, 2003, Goodman and Thibodeau, 2007). Bourassa et al. (1999) questioned the extent to which groupings of dwellings constructed from predefined geographical boundaries have the maximum degrees of internal homogeneity and external heterogeneity. Following the approach used in non-spatial housing market segmentation, we suggest that it is worth exploring further the idea of statistical specification of demand to form submarkets (Park 2013). Non-spatial approaches subdivide the market statistically but the divisions are inconsistence over time, since the market is constantly changing (Leishman, 2009, Jones et al., 2003). We propose an approach that subdivides spatial demand in a way that acknowlegdes the primacy of location in the house price calculus and does this using highly differentiated and spatial scale-specfic accessibility measures.

Kauko (2009) notes that the built environment is always an important factor in house price differentials due to its slow-changing nature. Noting this, our study builds upon the hypothesis that properties within a particular location tend to be similar in many attributes, and householders in a location tend to be from the same social-economic group and to share the same locational preferences (Ball and Kirwan, 1977). Furthermore, many studies within the built environment field have found that the accessibility information contained in a city's street layout is also associated with property value and by implication, locational preferences (Matthews and Turnbull, 2007, Xiao et al., 2014). In particular, street layout determines the ease of access to transport services, employment and all manner of human transactions within a city (Webster 2010, Jones et al. 2009). Jones et al. (2009) argue that particular urban forms imply particular inter-linkages of environmental, social and economic attributes, and that urban form tends to persist with submarket residential density. We extend these propositions about the importance of geography (Basu and Thibodeau, 1998, Goodman and Thibodeau, 2007, Watkins, 2001), by investigating the use of street segments as the basic spatial unit for aggregation and demarcation of spatial housing market demand.

The intuitive importance of streets and street segments (parts of streets between junctions) for housing homogeneity, identity and preferences lends additional credence to this method.

The aims of this paper are two fold: firstly, we address weaknesses in traditional attempts to identify spatial submarkets, and argue for the street segment as the basic spatial unit. Secondly, we test our idea empirically by comparing it with a traditional specification that requires local knowledge about building typology. We do this by clustering street segments on the basis of accessibility, creating stable, homogenous contiguous grouping of houses delineated by statistical tests that optimise internal homogeneity (Bourassa et al., 1999). We undertake this experiment using a case study of Cardiff, UK, the residential core of which was built over a short period of roughly 30 years at the turn of the 19th and 20th centuries.

This study follows a general framework for testing for sub-market existence at a given time as adopted in previous studies (Jones et al., 2009, Watkins, 2001, Leishman et al., 2013). To index street network segments for connectivity we utilize a spatial network analysis method known as ‘space syntax’(Chiaradia et al., 2012) and use two indices of connectivity, both measured at different spatial scales: “closeness” and “betweenness” (defined below). A two-step clustering analysis (Chen et al., 2009, Bacher et al., 2004) is applied to identify the optimal number of clusters necessary to delineate distinct housing submarkets. We use the indexed accessibility maps to identify sub-areas of the road grid that are characterized by similar accessibility value. These can be thought of as areas of the city characterized by homogenous values of local agglomeration economies since accessibility is a surrogate for all manner of urban agglomeration benefits and disbenefits. The remaining steps in our methodology follows Schnare and Struyk's (1976) approach, using the Chow test and a weighted standard error test to measure the efficiency of submarkets thus delineated.

The remainder of the paper is organized as follows. Section 2 outlines the theory of housing submarket specification and sets out our alternative approach based on street segment analysis. The methodologies of spatial network analysis and the specification of our spatial hedonic price model are reported in section 3. We introduce the study area and the dataset for the Cardiff in section 4. Results of three different submarket specification models (city-wide, building type, street layout) are presented in section 5; and the conclusions are presented in section 6.

2.0 Literature review

2.1 Specifications of housing submarkets

Housing submarkets are one cause of housing market disequilibrium according to (Whitehead and Odling-Smee, 1975). Straszheim (1975) first noted that a fundamental characteristic of urban housing markets is the variation in housing characteristics and price by location. Goodman and Thibodeau (1998) stated that a metropolitan housing area is always segmented into small submarkets due to a mix of supply and demandrelated factors. Therefore, understanding the segmentation of a housing market is essential for both policy makers and private property investment assessments， especially since housing submarkets can serve to reduce consumers' search and other transaction costs (Goodman and Thibodeau, 2007). However, there exists little consensus as to how spatial submarkets should be defined for applied housing studies (Adair et al., 1996, Watkins, 2001).

Three mainstream approaches exist for identifying submarkets. The first one is mainly focused on geographical areas (e.g. census tracts or local government areas) using predefined or otherwise convenient geographical boundaries to delineate the housing submarket. For example, Straszheim (1975) and Gabriel and Wolch (1984) use the racial composition of districts to delineate sub-markets; Sonsteilie and Portney (1980) use political districts; and Hancock (1991) uses postcode districts. Munro (1986) specified the Glasgow market using a pre-defined geographical boundary system: north and south of the river and inner and outer suburban areas. Additionally, researchers have emphasized that buildings' structural characteristics determine and reflect peoples' preferences on the supply-side, with willingness to pay for individual housing characteristics showing a consistency within particular building categories. For example, Dale-Johnson (1982) used a factor analysis to group dwellings with similar characteristics in order to define housing submarkets. Allen et al. (1995) specified housing submarkets on the basis of dwelling types (e.g., condominiums, single-family homes, and apartments).

Adair et al. (1996) attempted to subdivide the city into inner city, middle city and outer city and identified nine submarkets in the Belfast housing market, based on terraced, semidetached and detached dwellings within each area. Watkins (2001) used a hybrid definition that nests dwelling characteristic-based submarkets within a spatially defined market sub-structure. Goodman and Thibodeau (1998) defined a housing market as a geographic area, utilizing the hierarchical model to show how a metropolitan area can be segmented by school zone, as the premise of their assertions is that all homes within a spatially concentrated area share amenities directly associated with the property's location.

The second main approach to specify housing submarkets relies on data as a driver and emphasizes accuracy of estimation, allowing for systematic statistical methods (for example, principal component analysis and clustering) to delineate sub-markets. The regularities in consumer perceptions of individual characteristics of housing can be captured through multivariate statistical techniques in the same way as in the segmentation of any other consumer market. Bourassa et al. (1999), for example, segmented the Sydney and Melbourne housing markets by applying principal components and cluster analysis to a variety of neighborhood attributes, spatial and structure characteristics and lettings data in order to determine specific submarkets. Further, Day et al. (2003) used hierarchical clustering techniques to identify housing sub-markets defined by a combination of property types, locations and socioeconomic characteristics of residents. Tu et al (2007) investigated housing market segmentation through housing price spatial autocorrelation, seeking to let the data define urban housing market segmentation, rather than use the traditional administrative or other pre-defined boundaries to limit sub-market structure. Statistical delineation methods can, in principle, be used to cluster either individual housing units or to cluster individual areas.

A third approach to submarket definition emphasizes the self-reported specification on spatial boundary of housing submarket by expert, the sales agent an etc. This is exemplified by Palm (1978), who argued that existing methods that use fixed geographies are flawed, as she demonstrated that specifications based on brokers' evaluations are better than those based on economic and racial-ethnic characteristics of householders. Also re-examining the role of the supply-side broker, Bourassa et al. (2003) compared two submarket constructions: (1) geographically concentrated "sale areas" used by local real estate appraisers in New Zealand and (2) a spatial submarket construction obtained by applying a cluster analysis to the most influential factors generated from property, neighborhood and locational attributes. They concluded that while a statistically generated submarket significantly increased hedonic house price prediction accuracy, it did not outperform the sales-area submarket model.

Recent developments of statistical approaches to spatial sub-market identification include Chen et al. (2009), who conducted two clustering methods, namely the 'K mean' and ‘two-step’ methods, to compare housing market constructions in Knox County, Tennessee, USA. Their findings indicate that housing sub-market boundaries drawn from local government jurisdictions, school districts and expert opinions are closely aligned with the boundaries drawn by statistical clustering methods in a mature housing market.

However, Goodman and Thibodeau (2003, 2007) insist that researchers should impose submarket geographic boundaries rather than deriving them through statistical modeling, as practitioners and policy makers require a clear grasp of the housing submarket structure system across space. Furthermore, Jones et al. (2003) also point out that specifying housing submarkets by social neighborhood characteristics is unstable over time because the market is constantly changing. Such an approach requires constant monitoring of the spatial dynamics of am urban housing system since the sub-market boundaries are likely to change over time Leishman (2009). This makes emergent specifications of submarkets based on data difficult for policy purposes. Household self-selection as a process in sub-market formation is likely to be important in mature cities with well-defined social-spatial geographies and less so where the housing markets are immature.

2.2 An alternative submarket specification using street geometry and topology

It has long been accepted that accessibility is a major influence on residential location, with formal models going back to Von Thunen. Beckmann (1973) and Michelson (1977) noted that accessibility is jointly purchased with a residential plot of land. Theoretically accessibility can be viewed as an attribute of land. There are many studies that show accessibility to be empirically associated with property value. For example, Handy and Niemeir (1997) show that different segments of the population care about different sets of opportunities and evaluate the impedance to and the attractiveness of opportunities in distinct ways. Further, Niemeier (1997) showed that accessibility preference is connected with social neighborhood characteristics, finding that different social groups can be shown to value accessibility in disparate ways.

Kauko (2009) noted that the physical environment is always an important factor in house price differentials, while Jones et al. (2009) argued that street layout creates distinct urban structure determing how areas, space, place, and development sites are organized and therefore valued. Batty (2009) introduces the idea of ‘centrality’ as a type of accessibility that can be measured in a street network. Webster (2010) argues that network centrality in an urban configuration can be considered a city’s preeminent public good and that although there are sub-markets for it, there are strong and abiding patterns of demand for ‘raw’ geometric ‘general accessibility’. Empirical studies have confirmed statistically that street layout is associated with property value. For example, elements of the road network connectivity index such as intersection density and network “integration”, have been shown to positively impact housing prices (Xiao et al., 2014, Matthews and Turnbull, 2007). Furthermore, street layout network metrics have been found to capture the special accessibility demands of sub-groups as well as the general accessibility of a location's connectivity to all other locations. Vaughan et al. (2005) and Vaughan and Penn (2006), for example, show that poorer immigrants tend to congregate in poverty areas with lower road network connectivity.

The specific trigger of our study is to extend Goodman and Thibodeaus' (1998, 2003, 2007) and Bau and Thibodeaus' (1998) study of the spatial division of submarkets, but doing so using street segments as an alternative to geographic boundaries and aggregation units. The logic of our hypothesis is that the street segment (part of a street between two consecutive junctions) is a natural geographic boundary and unit in a city's subdivided housing market. It can be taken as a unit within which there is a high probability of homogenous households with respect to locational preferences for services, amenities, destination opportunities and the characteristics of neighbours. A sub-market segmentation scheme based on street segments has the advantages of retaining a consistent physical boundary (important for policy makers and market actors) while allowing greater precision of aggregation by virtue of the finer spatial scale and the greater behavioural importance of street segments compared to more arbitrary fixed data-collection boundaries.

3.0 Methodology

The study follows Schnare and Struyk's (1976) general framework for testing for sub-market existence at a given time, a framework widely employed in other studies such as Watkins (2001) and Bourassa et al. (1999). Generally, the procedure involves four stages. First, a hedonic price function is estimated for the entire market. Secondly, sub-markets are defined on two different dimensions: (i) building type; (ii) urban configurational features. Then, following Bourassa et al. (1999) and Chen et al (2009), we utilize a simple, two-step cluster analysis (Bacher et al. 2004) to identify the optimal number of groups of street segment features with similar accessibility values measured by the 'Space Syntax' method in Confeego 1.0. Thirdly, a Chow test is computed to establish whether statistically significant differences exist between the spatial sub-markets derived by the cluster analysis. Finally, a weighted standard error (WSE) is applied for the evaluation of sub-market sub-division. Normally, WSE represents the weighted average of the mean square error (MSE) of each submarket hedonic equations, and if the reduction of the weighted standard error is lower than the criteria value , then the postulated market subdivision is accepted (Goodman and Thibodeau, 2003, Park 2013).

3.1 Measuring accessibility in street layout.

Space Syntax, is a street network analysis method developed by UCL academics, (Hillier and Hanson, 1984). Space Syntax was designed principally to parameterize urban design and architectural schemes and borrows from more general network analysis. Contrasting with traditional geographic network analysis, space syntax models street segments as 'node' and intersections as 'link' (Batty and Rana, 2004).

In our study, we have used two space syntax metrics of network centrality: Integration and Route Choice. Behaviorally, these are based on two trip-choice criteria that an individual has to make while negotiating a road network (while using a home as a basis for making economic and social transactions in the city); namely, selecting a destination and selecting a route to get to the destination. The former variable is premised on how easy is it to get to a particular destination, and is termed the to-movement component. Destinations that are more accessible are more likely to be selected as locations by higher activity uses, such as shops. The latter variable determines the places that an individual has to pass through the get to a destination, this is termed the through-movement component.

In graph-theoretic terminology, the to-movement potential is termed 'closeness' (*integration* in space syntax) and measures the ease with which a destination may be accessed within a network. Space syntax closeness analysis models the mean distance between the origin and all possible destinations within network radius R. It measures the extent to which a road segment is close to all other road segments along the shortest distance of the street network and is formalised by equation (1):

$$Closeness_i = \frac{N-1}{\sum_{j=1;j\neq i}^{N} d_{ij}} \qquad \text{(Equation 1)}$$

Where N is the total number of segments in the network, and d_{ij} is the shortest topological depth between segment i and j.

The through-movement potential is captured by the graph-theoretic measure of betweenness (Freeman, 1977). This is commonly referred to as 'Route Choice' (or Choice) in the space syntax literature and measures the degree of potential for movement through a particular segment of the road network. In contrast to *'Closeness*,' which measures the relative ease of reaching potential destinations, the betweenness index indicates how often people are likely to pass through a particular route and therefore which parts of the road network will be the busiest. Space syntax betweenness analysis assumes that people will travel from two points on the network along the shortest path based on physical distance. As we are interested in road segments, a calculation of betweenness at radius R measures the number of shortest paths by distance connecting all pairs of road segments in the network with the maximum length of the path being R. A road segment is more central, and has more potential for through traffic the larger the number of shortest paths within the surrounding network that pass through it. Betweenness for road segment i is defined as:

$$betweenness_i = \frac{1}{(N-1)(N-2)} \sum_{j=1;k=1;j\neq k\neq i}^{N} \frac{n_{jk}(i)}{n_{jk}} \qquad \text{(Equation 2)}$$

Where n_{jk} is the number of shortest paths between segment j and k, and $n_{jk}(i)$ is the number of these shortest paths that contain segment i .

We measure closeness and betweenness at different radii: 400m 800m, 1200m, 1600m, 2000m, 2500m, 3000m, 4000m, 5000m, 6000m, 7000m, 8000m, 10000m and global Nm (the entire network) to capture various kinds of location advantages. We utilize these measurements as it is assumed that different householders assess a location for different kinds of centrality – access to city-wide destinations, local destinations, through traffic from the entire city, from the local area and so on. Each can be considered a potentially independent influence on location choice and therefore on price and market characteristic.

We implement these measurements using the Integrated Transport Network Layer (ITN) street network from the UK Ordnance Survey Mastermap, and the radii setting is determined by findings in the literature and bounded by size of the study area.

3.2 Hedonic model specification

The hedonic price model employed in this study is specified in the following general form:

$$P_i = \alpha_i + \beta_1 S_i + \beta_2 C_i + \beta_3 T_i + \beta_4 D_i + \beta_5 X_i + \varepsilon_i \qquad \text{(Equation 3)}$$

Where,

P_i= Transaction price of residential property;

S_i= Vector of property structural attributes;

C_i= Classification of OA;

T_i= Year of transacted property price;

D_i= Vector of conventional accessibility variables;

X_i= Vector of space syntax spatial network accessibility metrics at different radii;

ε_i = Random error term

As is common in hedonic property price research a log-linear (semi-log) specification is used (Malpezzi, 2003, Orford, 2000, Orford, 2002), as this can deal with dummy variables for characteristics that are either present or absent (0 or 1). It also reduces heteroscedasticity in the error terms (Diewert, 2003).

4.0 Data source and study area

In order to undertake the tests described above, Cardiff, the capital city of Wales, was chosen as a study area, due to the availability of hedonic data from Orford's (2000, 2010) previous studies. The study focuses on an area of 6x4 km stretching from the north of Cardiff city centre to the edge of the suburbs (Figure 1) containing a representative housing stock for the city, these include Victorian and Edwardian inner-city terraces, inter-war, post-war semi-detached and detached suburban homes; recent infill development of flats in the inner-city, and new builds on the urban edge. There also exists a dual carriage way (the A48M) dividing the study area into the inner-city and suburbs, with each displaying different social and built-form characteristics. This morphology is typical of many British cities.

[Insert Figure 1 Here]

The data sets are collected from several sources. Firstly, property prices come from the England and Wales Land Registry and a service license was acquired to use the following data: full address of property, price paid, sale date, property type (detached, semi-detached, terraced, Flat/Maisonette), new-build or not, and tenure (freehold or leasehold). Data for 16,297 properties sold in the study area during the period 2001 to 2007 (an average of 2000 transactions per year) were acquired and linked to Ordnance Survey Master map Address Layer that provided the grid co-ordinates for each property to a resolution of less than 1 meter. Since the Land Registry does not supply information on the size of property, floor area was estimated for each property in the property price database using a methodology described by Orford (2010) and the natural log of these measures was used in our hedonic models to standardize by size. Secondly, Office of National Statistics (ONS) Output Area Classification (OAC) data were used to capture area demographic and socio-economic characteristics (Vickers and Rees, 2006). We used seven OAC classes: blue collar communities, city living, countryside, prosperous suburbs, constrained by circumstances, typical traits and multicultural. Thirdly, land use information is from Ordnance Survey Master map. This data set includes location information of green spaces and hospitals and the ITN street network.

[Insert Table 1 Here]

Finally, 25 variables were prepared for the hedonic models: 21 dummy variables and four continuous variables, with an additional 28 continuous variables representing street layout features. Following housing price studies of Cardiff by Orford (2002, 2000), the natural log of distance from each property to the city centre is used in our study as a traditional central-accessibility attribute. Regarding structural characteristics, very few properties were new build and four-fifths were freehold tenure. Terraced houses made up the largest portion in the sample (53%), with semi-detached houses the second largest (21%), followed by flats (17%). Only 5% of the OAs in the study area were classified as 'constrained by circumstances' in contrast to 28% classified 'typical traits' and 'living in the city'. The percentage of total housing stock sold each year in the sample ranged from 11 to 15%, with the exception of 2008, when only 1% changed hands due to the large downturn in the housing market that occurred in that year. For dummy variables, we followed the rule of thumb, dropping the category of the lower marginal group (earliest year, cheapest housing class etc) as the reference variable to achieve intuitive expectations of the likely signs.

5.0 Empirical Results

5.1 Hedonic Results for the Complete Market

The initial step in the test procedure outlined above is to estimate a hedonic model for the city-wide market. This provides a reference estimation for comparing the efficiency of sub-market specifications. We employed robust regression estimation for all models in order to control the heteroscedasticity in residuals. Results of combined model are presented in Table 2, and they are broadly similar in performance to those reported elsewhere in the hedonic house price literature (Orford 1999, 2000, 2002). House price variance is explained by a range of structural, neighborhood, and locational characteristics. For example, building structure characteristics have a positive impact, whereas, locational attributes, such as the distance to CBD and distance to park show a 'trade-off' effect on property value. Generally, the results (Table 2) show that the models are statistically significant; the adjusted R–square of all models is 63.3%; and multicollinearity is not a problem with all VIF values under 10 for each variable.

[Insert Table 2 Here]

5.2 Aggregating street segment features

The second stage of our method identifies a potential housing submarket geography based on street segments, using the two-step cluster method of Bacher et al. (2004) to acquire an optimal numbers of clusters. It is noted that we clustered 28 attributes of street layout features measured at different radii around the homes for which we have price information, since, it represents the locational specific characteristics of all manner of urban agglomeration benefits and disbenefits. Table 3 displays the results showing that the optimal number of submarkets based on all street features is three, (since the critical ratio is above the critical value at 2.448). The total sample of 16297 is therefore divided into three subgroups, with subgroups respectively containing sample sizes of 8615, 4002 and 3680.

[Insert Table 3 Here]

[Insert Figure 2 Here]

With regards to the aggregation results, Table 4 and Figure 2 show that street layout features are segregated at different levels. This is reported as sub-group "1" which is associated with the areas near the central city, which has the greatest ease of access for pedestrian and automobile usage, and consequentially having the highest values of both *closeness* and *betweenness* at all radii level. Sub-group "3" captures the areas along secondary streets that are served by dead-end roads, and these are the areas with the second highest *betweenness* value and second lowest *closeness* value. In contrast, sub-group "2" has the lowest values of both *closeness* and *betweenness*, indicating that they are the most isolated area at both pedestrian and automobile level, where connectivity and potential traffic flows are both low.

[Insert Table 4 Here]

It is also apparent from Table 5 that street layout features can spatially discrimitate building structure and social characteristics of properties. For example, sub-group “1” comprises 52.9% of all houses in the study area and contains over 50% terraced houses, which also have the highest average housing price and floor area. More than 70% of residents are classified as ‘living in the city’ and share ‘typical traits’ characteristics. In contrast, just 24.6% of total observations are in sub-group “2”, which has the cheapest average housing price and smallest average floor area. 30% of residents are classified as coming from social economic groups drawn from the group of ‘typical traits’, and another third of the observations are from ‘blue collar’ and ‘multicultural’ groups. Subgroup “3” covers a sample size of 3680, mostly capturing Semi-detached houses with larger floor area. Two thirds here are from ‘blue collar’, ‘prosperous suburbs’ and ‘typical traits’ social classes. The univariate classification of street segments into distinct classes of accessibility is of interest in it own right. It shows that the city of Cardiff has three distinct classes of road links in terms of centrality. Further, we have shown that these classes are highly correlated with important attributes of housing markets.

[Insert Table 5 Here]

5.3 Hedonic estimations for alternative aggregations

In order to test submarket existence of our demarcation, we estimate for each submarket using the same attributes as in the city-wide model. It is important to note that the coefficients are relatively unimportant when testing for submarket existence (Dale-Johnson, 1982). However, the results shown in Table 5, indicate that it is indeed possible and useful to identify geographical areas that are statistically distinct in terms of price, network morphology and social morphology. Since, different socio-economic groups seem to be associated with particular patterns of accessibility (Table 6). For example, 'Constrained by Circumstances' group is insignificant in all subgroups, while the 'blue collar' group is only significant in sub-group '1.' The 'living in city' was a dominant determinant of distinct network characteristics in the city-wide model, but becomes less dominant or unimportant in submarket models, for example it appears along with 'typical traits' as explanatory of distinct accessibility in sub-group "3". Finally, The Chow test used to examine whether each subgroup differed, suggests there is no evidence of parameter equality between any submarkets (Table 7), thereby confirming that submarkets exit.

[Insert Table 6 Here]

[Insert Table 7 Here]

5.3. Comparison of Submarket Classifications

In order to understand the effectiveness of our submarket specification, we compare the results with existing traditional identifications (Watkins 2001, Bourassa et al , 1999, Chen et al 2009). In the local context of the study area and the UK more generally, there is a strong general cultural preference for detached over semi-detached, semi-over terraced and terraced over apartments. Orford (2000) confirms that these building structure features are important for housing submarkets in the Cardiff housing market. As the dataset in our study area has four types (detached house, semidetached house, flat and terrace house), the whole market is easily classified into four building-type submarkets, which we use to compare with our street segment-based approach

[Insert Table 8 Here]

Table 8 shows the estimation results for each submarket, revealing that the adjusted R square value varies from 58.4% to 65.7%. Only the estimation of the semidetached group is higher than the city–wide model (3% more). The general results are similar to our submarket specification using street segments, showing that building types clearly also reflect people's preferences, in particular for the two groups: 'blue collar' and 'Constrained by Circumstances'.

The Chow test similarly confirms that the four groups based on building typology are statistically not equal at a 1% confidence level (table 9). Thus, the results show that the delineation of housing submarkets can also (and more traditionally) be safely based upon building types.

[Insert Table 9 Here]

To compare the performance of these two submarket delineations, we take Schnare and

Struyk's (1976) `common-sense' test, comparing the reduction in the standard error (weighted) of the segmented model with the standard error of the market-wide hedonic model. Schnare and Struyk accepted a threshold of 10 percent reduction in their modelling, whereas Dale and Johnson (1982) suggested five percent. From Table 10, we can see that our street segment and building type submarkets both pass the weighted 10% standard error test, at 5.41% and 9.09% respectively. The building type sub-market, based on it local knowledge of cultural values has only 3.7% difference in performance compared to our alternative specification. This finding is supportive of Watkins (2001) and Orford's (2000) own systematic confirmations that in the UK context, dwelling type variables provide essential information for identifying housing submarkets.

Given the dominance of these cultural values, we expected this result. Although our scheme did not produce a superior sub-market estimation in comparison, it is important for a number of reasons as discussed in the next section. The results confirm that a submarket demarcation scheme based on street segments connectivity is valid one. We have identified three types of street segments which are able to discriminate housing price in a typical British Victorian city. We suggest that with further investigation in other contexts and refinement based on comparative findings, this scheme could comprise a more generic and adaptive methodology for spatial division of submarket in cities where there is a less well-established social valuation of housing typologies. It stands to reason that this is likely to be the case in the less mature housing markets of rapidly developing cities.

[Insert Table 10 Here]

6.0 Conclusions

It has long been established that urban housing markets are too complex and subdivided to be described adequately by a simple, unitary, competitive equilibrium model (Whitehead, 1999). Neither can they unambiguously be divided into submarkets on the basis of regularities in their heterogeneity. There are regularities but submarket patterns will depend on variables chosen and spatial and temporal scales used to measure them. This paper extends the search for alternative generic and specific schemes of subdivision. It does what has not been done before: to experiment with a subdivision based on street network accessibility (also referred to as connectivity or centrality). It hypothesizes that since accessibility is the key locational attribute of a residential home, then it should be possible to delineate housing submarkets on this basis. This requires both finely grained spatial units of analysis (street segments) and a reliable and accurate measure of accessibility (closeness and betweenness measures of network centrality). In this way, we have extended the analytical frameworks of Goodman and Thibodeau (1998, 2003, 2007), Sonsteilie and Portney (1980), Hancock (1991), and Watkins (2001), which emphasize that geographic areas should be the basis for delineating housing submarkets. By using a fine-grained spatial unit we were able to take a statistical approach to identify submarkets as emergent phenomenon and to do this in a way that is easily interpretable as spatial phenomenon. From this insight, we propose a novel idea: the street segments, as a type of natural geographic unit and boundary, is likely to be highly associated with people's preferences for locational attributes. Our results show that this is indeed the case and that the locational preferences of different groups are strongly correlated with the three classes of centrality identified in Cardiff's street pattern. The finding that the performance of our submarket segmentation model was only marginally lower than that using a priori knowledge about housing typology preferences confirms the worth of this line of enquiry.

There are three main contribution of this research. First, we have established that the street segment is a good surrogate of many kinds of spatially clustering housing attributes, and is a good spatial unit to capture centrality - something quite axiomatic and central to locational choice. The approach has potential for understanding the spatial structure of both polycentric and moncentric cities and cities with different housing typologies and with different social values attached to those typologies. As such, our approach less reliant on the data, prior knowledge and the existence of highly visible and culturally specific discriminating features of housing markets. It is a more generic approach that subdivides markets on a dimension that, from the historical literature as well as from intuition, is possibly of more abiding and a more generic importance than housing typology: centrality. Furthermore, such an approach combines the advantages of both having a geographically specific definition of submarkets and being subject to statistical verification in estimation (Goodman and Thibodeau, (2003, 2007).

Secondly, the results provide empirical evidence to support Jones and McDonald's (2004) assertion that urban form should be important in understanding housing market segregation. We have shown that optimal housing sub-markets based on street layout, are close to the specification based on *a priori* knowledge of building type (*a priori* information may be presumed to always improve submarket classification). Compared with previous studies using predefined geographic area, our method improves the resolution at street level, while it also has the potential to increase the accuracy of estimation (Bourassa et al. 1999).

Thirdly, we utilized spatial network metrics to measure the accessibility features of a street network (using Space Syntax software). The two accessibility indices: closeness and betweenness were effective in helping to delineate a spatially discrete pattern of street segment, which reinforces the findings reported by Maclennan and Tu (1996). Our findings also confirm that street layout is linked to spatial segregation (Jones et al. 2009, Vaughan 2007), since a street layout always spatially distributes access to services unequally. It is a novel idea that we can use this inequality to *a priori* identify social spatial regions within a city (and as we have done, use these to inform the identification of housing submarket).

Finally, to summarise, there are several features of our approach to submarket segmentation that suggests areas of further research that would be rewarding. First, the approach increases the resolution of the segmentation. The idea of micro-housing markets, or multi-scale submarkets is an intriguing one. Submarkets can be delineated at different scales of analysis to show the impact of different locational externalities. Second, our approach should be more useful for policy interventions than approaches using arbitrary territorial boundaries; all the more so if spatial submarkets can be identified at different nested spatial scales. Third, the approach can identify submarket boundaries in a way that approaches the accuracy of traditional models that require more specific and subjective knowledge requirements. Fourth, there is good deductive reason to suppose, and some evidence to confirm, that the approach may be suitable for delineating submarkets in cities with high building structure homogeneity (ie few visual cues from buildings themselves). Fifth, the analysis offers the possibility of assessing at the urban design and planning stage, the likely impact of urban grid configuration on housing submarket formation.

Acknowledgements. The authors would like to thank anonymous reviewers for their valuable comments and helpful suggestions. Alain Chiaradia and Chinmoy Sarkar from the sDNA group at Cardiff and Hong Kong made useful suggestions.

Appendix

1. The Chow test has the null hypothesis that $\beta_{01} = \beta_{02}, \beta_{11} = \beta_{12}, \beta_{k1} = \beta_{k2}$, and is formulated as follows:

$$F = \left[(RSS_c - RSS_i - RSS_j)/\mathrm{k}\right] / \left[(RSS_i + RSS_j)/(n + m - 2k)\right]$$
$$= \frac{[RSS_c - (RSS_i + RSS_j)](n + m - 2k)}{(RSS_i + RSS_j)k}$$

Where, n and m are the number of observations in the two sub-samples i and j; and RSS_c is the residual sum of squares of the combined model. k is the number of exploratory parameters, including the intercept. The RSS_s is found by estimating the equation three times, once for each of the sub-samples and once for the pooled sample. The F statistic which is calculated is compared with the critical value F_{cv} which is distributed as $F_{cv} \sim F(R + 1, n + m - 2(k + 1))$.

2. The formula for calculating the geometric mean of the standard error (SE) test of the j segmented models can be written as follow:

$$SE_c = \frac{N_1 - k_1 - 1}{\sum(N_j - k_j - 1)} SE_1 + \frac{N_2 - k_2 - 1}{\sum(N_j - k_j - 1)} SE_2 + \cdots + \frac{N_j - k_j - 1}{\sum(N_j - k_j - 1)} SE_j$$

Where, N_j is the number of transactions in the jth sub-market, k_j is the number of explanatory variables in the jth sub-market equation; and there are j sub-markets.

References:

ADAIR, A., BERRY, J. & MCGREAL, W. 1996. Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research,* 13**,** 67-83.

ALLEN, M. T., SPRINGER, T. M. & WALLER, N. G. 1995. Implicit pricing across residential rental submarkets. *The Journal of Real Estate Finance and Economics,* 11**,** 137-151.

BACHER, J., WENZIG, K. & VOGLER, M. 2004. *SPSS TwoStep Cluster-a first evaluation*, Lehrstuhl für Soziologie Berlin, DE.

BASU, S. & THIBODEAU, T. G. 1998. Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics,* 17**,** 61-85.

BATTY, M. & RANA, S. 2004. The automatic definition and generation of axial lines and axial maps. *Environment and Planning B,* 31**,** 615-640.

BOURASSA, S. C., HAMELINK, F., HOESLI, M. & MACGREGOR, B. D. 1999. Defining Housing Submarkets. *Journal of Housing Economics,* 8**,** 160-183.

BOURASSA, S. C., HOESLI, M. & PENG, V. S. 2003. Do housing submarkets really matter? *Journal of Housing Economics,* 12**,** 12-28.

BRAMLEY, G., LEISHMAN, C. & WATKINS, D. 2008. Understanding neighbourhood housing markets: Regional context, disequilibrium, sub-markets and supply. *Housing Studies,* 23**,** 179-212.

CHEN, Z., CHO, S.-H., POUDYAL, N. & ROBERTS, R. K. 2009. Forecasting housing prices under different market segmentation assumptions. *Urban Studies,* 46**,** 167-187.

CHIARADIA, A., HILLIER, B., SCHWANDER, C. & WEDDERBURN, M. 2012. Compositional and urban form effects on centres in Greater London. *Proceedings of the ICE-Urban Design and Planning,* 165**,** 21-42.

DALE-JOHNSON, D. 1982. An alternative approach to housing market segmentation using hedonic price data. *Journal of Urban Economics,* 11**,** 311-332.

DAY, B., SOCIAL, C. F. & ENVIRONMENT, E. R. O. T. G. 2003. *Submarket identification in property markets: a hedonic housing price model for Glasgow*, Centre for Social and Economic Research on the Global Environment.

DIEWERT, W. E. Hedonic regressions: a review of some unresolved issues. 7th Meeting of the Ottawa Group, Paris, May, 2003. 29.

GABRIEL, S. A. & WOLCH, J. R. 1984. Spillover effects of human service facilities in a racially segmented housing market. *Journal of Urban Economics,* 16**,** 339-350.

GOODMAN, A. C. & THIBODEAU, T. G. 1998. Housing Market Segmentation. *Journal of Housing Economics,* 7**,** 121-143.

GOODMAN, A. C. & THIBODEAU, T. G. 2003. Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics,* 12**,** 181-201.

GOODMAN, A. C. & THIBODEAU, T. G. 2007. The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics,* 35**,** 209-232.

HANCOCK, K. 1991. The determination of housing submarkets: case studies using Scottish data. *unpublished paper, Centre for Housing Research, University of*

Glasgow, Glasgow.
HANDY, S. L. & NIEMEIER, D. A. 1997. Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A,* 29**,** 1175-1194.
HILLIER, B. & HANSON, J. 1984. *The social logic of space*, Cambridge University Press Cambridge.
JONES, C., LEISHMAN, C. & MACDONALD, C. 2009. Sustainable urban form and residential development viability. *Environment and planning. A,* 41**,** 1667.
JONES, C., LEISHMAN, C. & WATKINS, C. 2003. Structural change in a local urban housing market. *Environment and Planning A,* 35**,** 1315-1326.
KAUKO, T. 2009. Classification of Residential Areas in the Three Largest Dutch Cities Using Multidimensional Data. *Urban Studies,* 46**,** 1639-1663.
LEISHMAN, C. 2001. House building and product differentiation: An hedonic price approach. *Journal of Housing and the Built Environment,* 16**,** 131-152.
LEISHMAN, C. 2009. Spatial change and the structure of urban housing sub-markets. *Housing Studies,* 24**,** 563-585.
LEISHMAN, C., COSTELLO, G., ROWLEY, S. & WATKINS, C. 2013. The predictive performance of multilevel models of housing sub-markets: A comparative analysis. *Urban Studies,* 50**,** 1201-1220.
MACLENNAN, D. & TU, Y. 1996. Economic perspectives on the structure of local housing systems. *Housing studies,* 11**,** 387-406.
MALPEZZI, S. 2003. Hedonic pricing models: a selective and applied review. *Section in Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan*.
MATTHEWS, J. W. & TURNBULL, G. K. 2007. Neighborhood street layout and property value: the interaction of accessibility and land use mix. *The Journal of Real Estate Finance and Economics,* 35**,** 111-141.
MUNRO, M. 1986. Testing for segmentation in the private housing market in Glasgow. *Centre for Housing Research, Discussion Paper No,* 8.
NIEMEIER, D. A. 1997. Accessibility: an evaluation using consumer welfare. *Transportation,* 24**,** 377-396.
ORFORD, S. 2000. Modelling spatial structures in local housing market dynamics: a multilevel perspective. *Urban Studies,* 37**,** 1643.
ORFORD, S. 2002. Valuing locational externalities: a GIS and multilevel modelling approach. *Environment and Planning B,* 29**,** 105-128.
ORFORD, S. 2010. Towards a data-rich infrastructure for housing-market research: deriving floor-area estimates for individual properties from secondary data sources. *Environment and planning. B, Planning & design,* 37**,** 248.
PALM, R. 1978. Spatial segmentation of the urban housing market. *Economic Geography***,** 210-221.
PARK, J. 2013. The division of spatial housing submarkets: a theory and the case of Seoul. *Environment and Planning A,* 45**,** 668-690.
POUDYAL, N. C., HODGES, D. G. & MERRETT, C. D. 2009. A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy,* 26**,** 975-983.

PRYCE, G. 2013. Housing submarkets and the lattice of substitution. *Urban Studies,* 50, 2682-2699.
SCHNARE, A. B. & STRUYK, R. J. 1976. Segmentation in urban housing markets. *Journal of Urban Economics,* 3, 146-166.
SONSTELIE, J. C. & PORTNEY, P. R. 1980. Gross rents and market values: Testing the implications of Tiebout's hypothesis* 1. *Journal of Urban Economics,* 7, 102-118.
STRASZHEIM, M. R. 1975. An Econometric Analysis of the Urban Housing Market. *NBER Books*.
TU, Y., SUN, H. & YU, S. M. 2007. Spatial autocorrelations and urban housing market segmentation. *The Journal of Real Estate Finance and Economics,* 34, 385-406.
VAUGHAN, L., CLARK, D. L. C., SAHBAZ, O. & HAKLAY, M. M. 2005. Space and exclusion: does urban morphology play a part in social deprivation? *Area,* 37, 402-412.
VAUGHAN, L. & PENN, A. 2006. Jewish immigrant settlement patterns in Manchester and Leeds 1881. *Urban Studies,* 43, 653.
VICKERS, D. & REES, P. 2006. Introducing the area classification of output areas. *POPULATION TRENDS-LONDON-,* 125, 15.
WATKINS, C. A. 2001. The definition and identification of housing submarkets. *Environment and Planning A,* 33, 2235-2254.
WEBSTER, C. 2010. Pricing accessibility: Urban morphology, design and missing markets. *Progress in Planning,* 73, 77-111.
WHITEHEAD, C. M. E. 1999. Urban housing markets: theory and policy. *Handbook of regional and urban economics,* 3, 1559-1594.
WHITEHEAD, C. M. E. & ODLING-SMEE, J. 1975. Long-run equilibrium in urban housing-a note. *Urban Studies,* 12, 315.
XIAO, Y., WEBSTER, C. & ORFORD, S. 2014. Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. *Urban Studies*.

Figures:

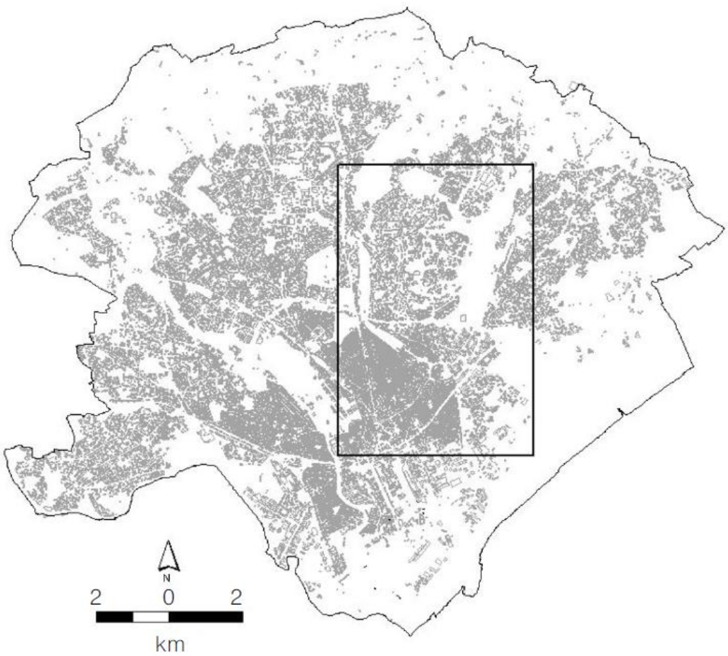

Figure 1 Study area in Cardiff, Wales, UK.

Source: Orford (2010)

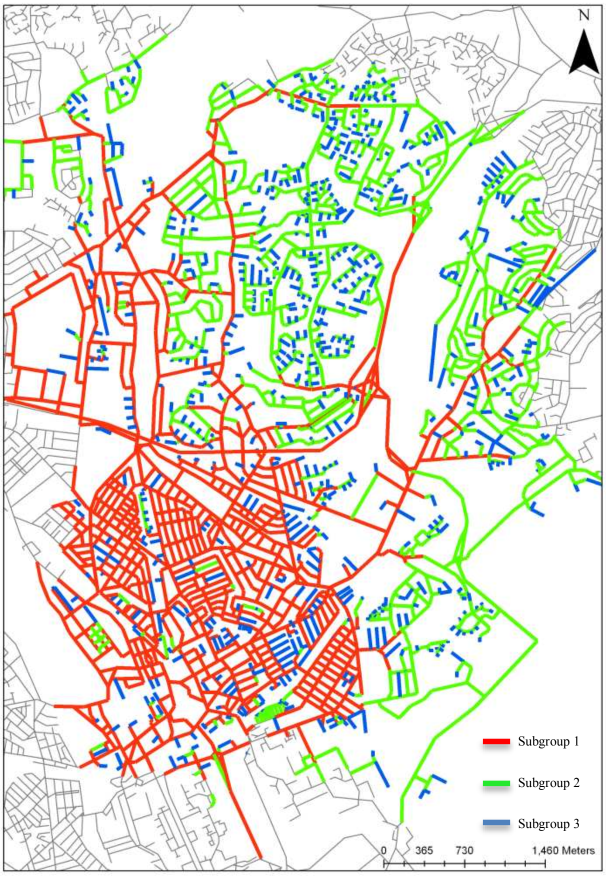

Figure 2 The pattern of street segments clustered by network centrality measures

Source: Authors

Tables:

Table 1 Description of variables

Variables	Description	Min	Max	Mean	SDev
LN_FL	Natural log of floor area	1.793	9.595	4.903	1.071
LN_CBD	Natural log of distance to City Centre	-0.66	1.79	0.76	0.62
LN_BAY	Natural log of distance to Cardiff bay	0.93	2.18	1.6	0.29
LN_ROATH	Natural log of distance to Roath Park	-3.19	1.52	0.72	0.52
LN_HOSP	Natural log of distance to Heath hospital	-0.42	1.82	0.99	0.48
LN_BUTE	Natural log of distance to Bute Park	-1.03	1.64	0.95	0.39
BE_R400M	Betweeness value at radius 400m	0	3.183	1.483	0.923
BE_R800M	Betweeness value at radius 800m	0	3.879	2.152	1.277
BE_R1200M	Betweeness value at radius 1200m	0	4.464	2.51	1.485
BE_R1600M	Betweeness value at radius 1600m	0	4.961	2.745	1.626
BE_R2000M	Betweeness value at radius 2000m	0	5.287	2.912	1.721
BE_R2500M	Betweeness value at radius 2500m	0	5.57	3.069	1.817
BE_R3000M	Betweeness value at radius 3000m	0	5.789	3.195	1.891
BE_R4000M	Betweeness value at radius 4000m	0	6.155	3.389	2.004
BE_R5000M	Betweeness value at radius 5000m	0	6.43	3.529	2.091
BE_R6000M	Betweeness value at radius 6000m	0	6.655	3.634	2.161
BE_R7000M	Betweeness value at radius 7000m	0	6.887	3.719	2.209
BE_R8000M	Betweeness value at radius 8000m	0	7.058	3.778	2.249
BE_R10000M	Betweeness value at radius 10000m	0	7.273	3.834	2.295
BE_N	Betweeness value for whole city	0	7.567	3.834	2.337
CL_R400M	Closeness value at radius 400m	0	95.898	29.479	16.135
CL_R800M	Closeness value at radius 800m	0	194.551	68.402	40.154
CL_R1200M	Closeness value at radius 1200m	11.928	350.988	120.269	67.933
CL_R1600M	Closeness value at radius 1600m	16.489	482.002	181.727	97.358
CL_R2000M	Closeness value at radius 2000m	26.903	576.524	246.244	125.338
CL_R2500M	Closeness value at radius 2500m	30.971	719.865	327.692	155.646
CL_R3000M	Closeness value at radius 3000m	44.303	825.042	411.021	181.137
CL_R4000M	Closeness value at radius 4000m	77.035	1044.48	584.539	219.047
CL_R5000M	Closeness value at radius 5000m	141.572	1317.72	763.448	254.955
CL_R6000M	Closeness value at radius 6000m	251.813	1604.22	944.528	281.435
CL_R7000M	Closeness value at radius 7000m	359.996	1793.71	1114.311	298.407
CL_R8000M	Closeness value at radius 8000m	441.707	1939.79	1248.791	308.14
CL_R10000M	Closeness value at radius 10000m	616.409	2107.21	1412.711	292.122
CL_N	Closeness value for whole city	858.122	2150.76	1521.65	245.688

Variables	Description	Type	Code 0 (%)	Code 1 (%)	Mean	SDev
DU_NEW	New Build	Dummy	92.9	7.1	0.08	0.268
DU_DET	Detached House	Dummy	91.2	8.8	0.1	0.3
DU_SEMI	Semidetached House	Dummy	79.2	20.8	0.21	0.407
DU_TER	Terrace house	Dummy	46.6	53.4	0.52	0.5
DU_FLAT	Flat	Dummy	83	17	0.17	0.375
DU_TEN	Tenure (Freehold=1 Leasehold =0)	Dummy	21.7	78.3	0.79	0.411
DU_BC	OAC Blue collar communities	Dummy	89.1	10.9	0.11	0.31

DU_CL	OAC Living in the city	Dummy	72	28	0.27	0.446
DU_PS	OAC Prosperous suburbs	Dummy	87	13	0.15	0.354
DU_CC	OAC Constrained by Circumstances	Dummy	95.5	4.5	0.05	0.208
DU_TT	OAC Typical traits	Dummy	71.7	28.3	0.28	0.449
DU_MU	OAC Multicultural	Dummy	84.7	15.3	0.15	0.353
Y2000	Transactions in 2000	Dummy	89.3	10.7	0.11	0.309
Y2001	Transactions in 2001	Dummy	86.6	13.4	0.13	0.34
Y2002	Transactions in 2002	Dummy	85	15	0.15	0.359
Y2003	Transactions in 2003	Dummy	86.7	13.3	0.13	0.339
Y2004	Transactions in 2004	Dummy	87.5	12.5	0.13	0.332
Y2005	Transactions in 2005	Dummy	90.2	9.8	0.1	0.301
Y2006	Transactions in 2006	Dummy	87.6	12.4	0.12	0.328
Y2007	Transactions in 2007	Dummy	88.2	11.8	0.12	0.321
Y2008	Transactions in 2008	Dummy	98.9	1.1	0.01	0.099

Table 2 Hedonic estimation results

	Coef.	Robust Std. Err.	t	Sig.	VIF
LN_FLOOR	0.13	0.00	33.70	0.00**	2.37
DU_NEW	0.20	0.01	16.20	0.00**	1.29
DU_DET	0.70	0.02	40.83	0.00**	3.19
DU_SEMI	0.43	0.02	28.65	0.00**	5.13
DU_TER	0.25	0.01	17.44	0.00**	7.19
DU_TEN	0.26	0.01	21.19	0.00**	3.45
DU_BC	-0.10	0.01	-7.97	0.00**	2.02
DU_CL	0.13	0.01	13.38	0.00**	2.47
DU_PS	0.38	0.01	29.04	0.00**	2.67
DU_CC	-0.05	0.02	-2.86	0.00**	1.48
DU_TT	0.13	0.01	14.13	0.00**	2.38
Y2001	0.12	0.01	10.36	0.00**	1.95
Y2002	0.30	0.01	27.20	0.00**	2.03
Y2003	0.51	0.01	45.72	0.00**	1.95
Y2004	0.68	0.01	59.79	0.00**	1.91
Y2005	0.78	0.01	64.83	0.00**	1.76
Y2006	0.81	0.01	71.31	0.00**	1.90
Y2007	0.85	0.01	74.94	0.00**	1.93
Y2008	0.84	0.03	30.28	0.00**	1.09
LN_DCBD	-0.10	0.01	-13.51	0.00**	2.70
LN_DRoath	-0.18	0.01	-14.67	0.00**	5.44
LN_DHeath	-0.07	0.02	-4.26	0.00**	6.07
_cons	10.19	0.03	352.76	0.00**	
Adj. R Square			0.633		
F-test			1209.32		
P-value			0．00		
RSS			1943.44		
Stand Error			0. 119419		

Significance:*<0.05 **<0.01

Table 3 Cluster results on street segments

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Changea	Ratio of BIC Changesb	Ratio of Distance Measuresc
1	316823.28			
2	172256.88	-144566.40	1.00	2.49
3	114410.82	-57846.06	0.40	2.45
4	91102.34	-23308.48	0.16	1.48
5	75539.26	-15563.08	0.11	1.72
6	66699.48	-8839.78	0.06	1.48
7	60902.69	-5796.79	0.04	1.07
8	55544.04	-5358.66	0.04	1.68

Table 4 Description of street features in each subgroup

	Subgroup 1		Subgroup 2		Subgroup 3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
BE_R400M	0.07	0.27	1.75	0.51	2.15	0.42
CL_R400M	15.91	8.95	24.54	10.50	45.55	13.10
BE_R800M	0.05	0.24	2.53	0.50	3.17	0.37
CL_R800M	39.38	27.38	52.60	24.95	109.84	29.29
BE_R1200M	0.06	0.27	2.94	0.54	3.71	0.42
CL_R1200M	74.54	50.95	91.46	42.31	192.16	42.98
BE_R1600M	0.07	0.29	3.21	0.59	4.06	0.47
CL_R1600M	117.21	75.27	140.48	64.71	285.01	53.16
BE_R2000M	0.08	0.33	3.41	0.62	4.30	0.52
CL_R2000M	165.02	99.71	193.07	85.77	378.65	62.45
BE_R2500M	0.08	0.34	3.60	0.65	4.53	0.57
CL_R2500M	228.68	128.45	261.59	107.46	491.71	68.71
BE_R3000M	0.09	0.35	3.76	0.67	4.70	0.62
CL_R3000M	296.79	153.24	336.86	128.37	599.33	73.96
BE_R4000M	0.10	0.40	4.01	0.72	4.95	0.70
CL_R4000M	445.68	197.09	506.70	166.60	799.50	79.73
BE_R5000M	0.10	0.39	4.18	0.76	5.15	0.73
CL_R5000M	600.94	239.34	678.58	196.28	1007.08	88.01
BE_R6000M	0.09	0.40	4.32	0.79	5.29	0.77
CL_R6000M	763.43	262.77	847.98	210.38	1216.85	97.80
BE_R7000M	0.10	0.42	4.42	0.82	5.41	0.80
CL_R7000M	918.37	273.59	1009.93	218.79	1408.54	104.06
BE_R8000M	0.10	0.41	4.49	0.85	5.49	0.83
CL_R8000M	1043.39	281.19	1143.76	230.91	1550.48	104.00
BE_R10000M	0.10	0.40	4.56	0.91	5.56	0.88
CL_R10000	1217.10	272.17	1321.03	222.69	1691.42	94.45
BE_N	0.03	0.19	4.58	1.00	5.58	0.89
CL_N	1341.88	223.71	1445.13	182.86	1757.27	92.78

Number	3879

Table 5 Description of each subgroup

	Subgroup 1			Subgroup 2			Subgroup 3		
N	8615			4002			3680		
% of Total	52.9			24.6			22.6		
	N	Percent	Mean	N	Percent	Mean	N	Percent	Mean
PRICE	8615	100.00%	148772.92	3720	100.00%	124949.99	3680	100.00%	127166.22
FLOOR_AREA	8615	100.00%	895.93	3720	100.00%	128.26	3680	100.00%	193.24
DU_NEW	573	6.70%	0.07	219	5.50%	0.05	240	6.50%	0.07
DU_DE	545	6.30%	0.06	345	8.60%	0.09	538	14.60%	0.15
DU_SEMI	1104	12.80%	0.13	1056	26.40%	0.26	1248	33.90%	0.34
DU_TER	5160	59.90%	0.60	2155	53.80%	0.54	1379	37.50%	0.37
DU_FLAT	1806	21.00%	0.21	446	11.10%	0.11	515	14.00%	0.14
DU_TEN	6616	76.80%	0.77	3295	82.30%	0.82	2845	77.30%	0.77
DU_BC	148	1.70%	0.02	760	19.00%	0.19	874	23.80%	0.24
DU_CL	3715	43.10%	0.43	483	12.10%	0.12	369	10.00%	0.10
DU_PS	743	8.60%	0.09	566	14.10%	0.14	804	21.80%	0.22
DU_CC	84	1.00%	0.01	302	7.50%	0.08	349	9.50%	0.09
DU_TT	2410	28.00%	0.28	1207	30.20%	0.30	999	27.10%	0.27
DU_MU	1515	17.60%	0.18	684	17.10%	0.17	285	7.70%	0.08
Y2000	887	10.30%	0.10	436	10.90%	0.11	421	11.40%	0.11
Y2001	1197	13.90%	0.14	505	12.60%	0.13	479	13.00%	0.13
Y2002	1299	15.10%	0.15	585	14.60%	0.15	501	13.60%	0.14
Y2003	1081	12.50%	0.13	544	13.60%	0.14	511	13.90%	0.14
Y2004	996	11.60%	0.12	527	13.20%	0.13	490	13.30%	0.13
Y2005	838	9.70%	0.10	427	10.70%	0.11	356	9.70%	0.10
Y2006	1111	12.90%	0.13	457	11.40%	0.11	452	12.30%	0.12
Y2007	1114	12.90%	0.13	475	11.90%	0.12	438	11.90%	0.12
Y2008	92	1.10%	0.01	46	1.10%	0.01	32	0.90%	0.01
LN_DCBD	8615	100.00%	0.42	4002	100.00%	1.01	3680	100.00%	1.26
LN_DROATH	8615	100.00%	0.66	4002	100.00%	0.75	3680	100.00%	0.83
LN_DHEALTH	8615	100.00%	0.83	4002	100.00%	1.04	3680	100.00%	1.14

Table 6 Estimations of each subgroup

	subgroup 1		subgroup 2		subgroup 3	
	Coeff.	t	Coeff.	t	Coeff.	t
LN_FLOOR	0.18	30.75**	0.15	11.30**	0.02	2.59**
DU_NEW	0.19	8.99**	0.14	5.84**	0.29	11.94**
DU_DET	0.60	17.45**	0.68	19.00**	0.73	21.45**
DU_SEMI	0.38	12.22**	0.44	13.53**	0.37	13.18**
DU_TER	0.19	6.62**	0.27	8.3**	0.17	6.12**
DU_TEN	0.41	15.33**	0.13	5.85**	0.14	5.95**
DU_BC	-0.07	-2.56*	-0.03	-1.36	-0.01	-0.62
DU_CL	0.14	11.29**	0.16	7.85**	0.09	2.94**

DU_PS	0.25	11.61**	0.42	16.15**	0.46	17.76**
DU_CC	-0.06	-1.18	-0.01	-0.49	0.01	0.55
DU_TT	0.08	6.77**	0.19	9.79**	0.17	8.29**
Y2001	0.11	7.24**	0.09	4.33**	0.15	6.37**
Y2002	0.31	18.98**	0.25	11.95**	0.30	12.55**
Y2003	0.50	30.69**	0.51	25.55**	0.54	23.40**
Y2004	0.69	42.49**	0.64	32.08**	0.70	29.89**
Y2005	0.78	47.60**	0.76	39.73**	0.83	36.32**
Y2006	0.80	53.87**	0.80	42.91**	0.85	39.92**
Y2007	0.83	54.54**	0.85	45.32**	0.91	41.93**
Y2008	0.82	27.03**	0.79	15.69**	0.97	28.12**
LN_DCBD	0.03	2.12*	-0.15	-9.78**	-0.02	-1.15
LN_DRoath	-0.31	-10.82**	-0.13	-6.94**	0.25	4.49**
LN_DHeath	0.12	3.82**	-0.13	-4.65**	-0.64	-7.15**
_cons	9.78	232.84**	10.25	131.76**	10.90	146.5**
Adj. R Square	0.592		0.679		0.721	
F-test	566.71		401.21		356.82	
P-value	0.00		0.00		0.00	
RSS	1088.528		376.570		368.047	
Stand Error	0.126691		0.094639		0.100642	

Significance:*<0.05 **<0.01

Table 7 Chow test for street features specification

Segments	Chow
Subgroup 1 with Subgroup 2	178.45**
Subgroup 1 with Subgroup 3	178.01**
Subgroup 2 with Subgroup 3	534.52**
** indicates significance at 1% level	

Note: in order to test for the differences in housing prices between submarkets, we used a Chow test for structural instability (parameter constancy) over space (more information on the Chow test is found in the appendix).

Table 8 Estimations results for building type specification

Dwelling type	N	Adjust-R square	F-statistics	RSS	Stand Error	Significant variables	Numbers of variables
Flat	2767	0.630	253.37	324.451	0.118111	LN_FLOOR,DU_NEW,DU_TEN, DU_CL, DU_PS, DU_TT, Y2001, Y2002, Y2003,Y2004,Y2005,Y2006,Y2007,Y2008, LN_DCBD,LN_DHealth	16
Detached	1428	0.625	117.31	170.720	0.121250	LN_FLOOR, , DU_NEW, DU_TEN, DU_CL, DU_PS, DU_CC, DU_TT, Y2001, Y2002, Y2003, Y2004, Y2005, Y2006,Y2007, Y2008, LN_DRoath, LN_DHealth	15

Semidetached	3408	0.657	326.91	316.882	0.093531	LN_FLOOR,DU_NEW,DU_TEN, DU_BC, DU_CL, DU_PS, DU_TT, Y2000, Y2001, Y2002, Y2003,Y2004,Y2005,Y2006,Y2007,Y2008, LN_DCBD, LN_Droath, LN_DHealth	18
Terraced	8694	0.584	624.43	948.562	0.109357	LN_FLOOR,,DU_TEN, DU_BC, DU_CL, DU_PS, DU_CC, DU_TT, Y2001, Y2002,Y2003,Y2004,Y2005,Y2006,Y2007,Y2008, LN_DCBD, LN_Droath, LN_DHealth	18

Table 9 Chow test for building type specification

Segments	Chow
Flat with detached	607.6248**
Flat with semidetached	608.5278**
Flat with terraced	300.7417**
Detached with semidetached	696.7662**
Detached with terraced	371.1832**
Semidetached with terraced	315.8658**
Note: ** indicates significance at 1% level	

Table 10 Weighted standard error tests.

Stratification scheme	Standard error	% reduction
Market-wide model	0.119419	
Structure definition identified by dwelling type	0.108566	9.09%
Submarket specified by urban configurational features	0.112962	5.41%

Note: The "common sense test", weighted standard error test is employed to compare the fitness of different submarket classification. This indicates how closely substitutable the housing units are in that market segment with those in other segments. More information is found in the appendix.